(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
8 July 2004 (08.07.2004)

PCT

(10) International Publication Number
WO 2004/057493 A2

(51) International Patent Classification[7]: G06F 17/30

(21) International Application Number:
PCT/GB2003/005096

(22) International Filing Date:
24 November 2003 (24.11.2003)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
0229625.9 19 December 2002 (19.12.2002) GB

(71) Applicant *(for all designated States except US)*: **BRITISH TELECOMMUNICATIONS PUBLIC LIMITED COMPANY** [GB/GB]; BT Group Legal, Intellectual Property Department, PP C5A ,BT Centre, 81 Newgate Street, London EC1A 7AJ (GB).

(72) Inventor; and

(75) Inventor/Applicant *(for US only)*: **STENTIFORD, Frederick, Warwick, Michael** [GB/GB]; Sheepstor, Boyton, Woodbridge, Suffolk IP12 3LH (GB).

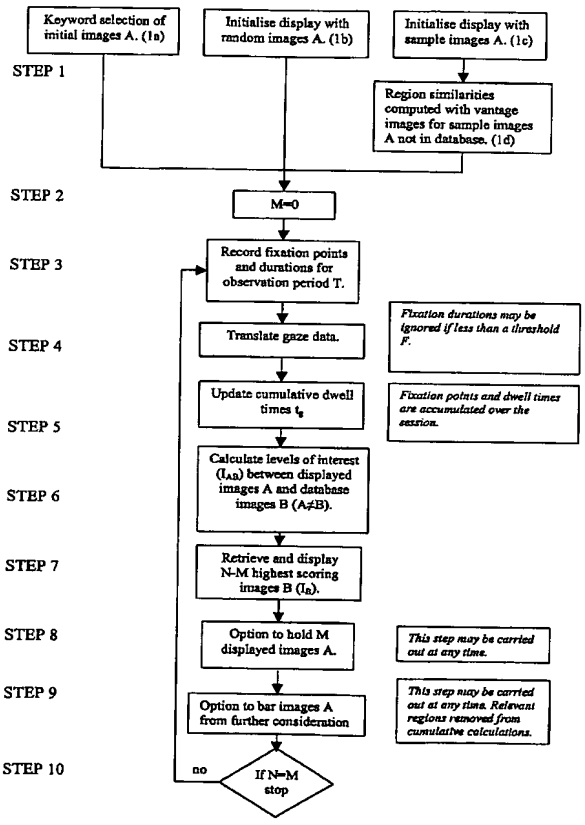
(74) Agent: **LLOYD, Barry, George, William**; BT Group Legal Intellectual Property Department, Holborn Centre, 8th Floor, 120 Holborn, London EC1N 2TE (GB).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (BW, GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW),

*[Continued on next page]*

(54) Title: SEARCHING IMAGES




(57) Abstract: A database of visual images includes metadata having, for a particular image, at least one entry specifying: a part of that image, another stored image, and a measure $S_{abi}$ of the degree of similarity between that specified part and the specified other image. The searching method comprises displaying one or more images; receiving input from a user (for example by using a gaze tracker) indicative of part of the displayed images; determining measures of interest for each of a plurality of non-displayed stored images specified by the metadata for the displayed image(s), as a function of the similarity measure(s) and the relationship between the user input and the part specified; and, on the basis of these measures, selecting, from those non-displayed stored images, further images for display.

WO 2004/057493 A2

Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

**Published:**

— *without international search report and to be republished upon receipt of that report*

*For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*

# SEARCHING IMAGES

The wide availability of digital sensor technology together with the falling price of storage devices has spurred an exponential growth in the volume of image material being captured for a range of applications. Digital image collections are rapidly increasing in size and include basic home photos, image based catalogues, trade marks, fingerprints, mugshots, medical images, digital museums, and many art and scientific collections. It is not surprising that a great deal of research effort over the last five years has been directed at developing efficient methods for browsing, searching and retrieving images [1,2].

Content-based image retrieval requires that visual material be annotated in such a way that users can retrieve the images they want efficiently and effortlessly. Current systems rely heavily upon textual tagging and measures (eg colour histograms) that do not reflect the image semantics. This means that users must be very conversant with the image features being employed by the retrieval system in order to obtain sensible results and are forced to use potentially slow and unnatural interfaces when dealing with large image databases. Both these barriers not only prevent the user from exploring the image set with high recall and precision rates, but the process is slow and places a great burden on the user.

**Prior Art**

Early retrieval systems made use of textual annotation [3] but these approaches do not always suit retrieval from large databases because of the cost of the manual labour involved and the inconsistent descriptions, which by their nature are heavily dependent upon the individual subjective interpretation placed upon the material by the human annotator. To combat these problems techniques have been developed for image indexing that are based on their visual content rather than highly variable linguistic descriptions.

It is the job of an image retrieval system to produce images that a user wants. In response to a user's query the system must offer images that are similar in some user-defined sense. This goal is met by selecting features thought to be important in human visual perception and using them to measure relevance to the query. Colour, texture, local shape and layout in a variety of forms are the most widely used features in image retrieval [4,5,6,7,8,9,10]. One of the first commercial image search engines was QBIC [4] which executes user queries against a database of pre-extracted features. VisualSEEk [7] and SaFe [11] determine similarity by measuring image regions using both colour parameters and spatial relationships and obtain better performance than histogramming methods that use colour information alone. NeTra [8] also relies upon image segmentation to carry out region-based searches that allow the user to select example regions and lay emphasis on image attributes to focus the search. Region-based

querying is also favoured in Blobworld [6] where global histograms are shown to perform comparatively poorly on images containing distinctive objects. Similar conclusions were obtained in comparisons with the SIMPLIcity system [30]. The Photobook system [5] endeavours to use compressed representations that preserve essential similarities and are "perceptually complete". Methods for measuring appearance, shape and texture are presented for image database search, but the authors point out that multiple labels can be justifiably assigned to overlapping image regions using varied notions of similarity.

Analytical segmentation techniques are sometimes seen as a way of decomposing images into regions of interest and semantically useful structures [21-23,45]. However, object segmentation for broad domains of general images is difficult, and a weaker form of segmentation that identifies salient point sets may be more fruitful [1].

Relevance feedback is often proposed as a technique for overcoming many of the problems faced by fully automatic systems by allowing the user to interact with the computer to improve retrieval performance [31,43]. In Quicklook [41] and ImageRover [42] items identified by the user as relevant are used to adjust the weights assigned to the similarity function to obtain better search performance. More information is provided to the systems by the users who have to make decisions in terms specified by the machine. MetaSeek maintains a performance database of four different online image search engines and directs new queries to the best performing engine for that task [40]. PicHunter [12] has implemented a probabilistic relevance feedback mechanism that predicts the target image based upon the content of the images already selected by the user during the search. This reduces the burden on unskilled users to set quantitative pictorial search parameters or to select images that come closest to meeting their goals. Most notably the combined use of hidden semantic links between images improved the system performance for target image searching. However, the relevance feedback approach requires the user to reformulate his visual interests in ways that he frequently does not understand.

Region-based approaches are being pursued with some success using a range of techniques. The SIMPLIcity system [30] defines an integrated region matching process which weights regions with 'significance credit' in accordance with an estimate of their importance to the matching process. This estimate is related to the size of the region being matched and whether it is located in the centre of the image and will tend to emphasise neighbourhoods that satisfy these criteria. Good image discrimination is obtained with features derived from salient colour boundaries using multimodal neighbourhood signatures [13-15,36]. Measures of colour coherence [16,29] within small neighbourhoods are employed to incorporate some spatial information when comparing images. These methods are being deployed in the 5$^{th}$ Framework

project ARTISTE [17, 18, 20] aimed at automating the indexing and retrieval of the multimedia assets of European museums and Galleries. The MAVIS-2 project [19] uses quad trees and a simple grid to obtain spatial matching between image regions.

Much of the work in this field is guided by the need to implement perceptually based systems that emulate human vision and make the same similarity judgements as people. Texture and colour features together with rules for their use have been defined on the basis of subjective testing and applied to retrieval problems [24]. At the same time research into computational perception is being applied to problems in image search [25,26]. Models of human visual attention are used to generate image saliency maps that identify important or anomalous objects in visual scenes [25,44]. Strategies for directing attention using fixed colour and corner measurements are devised to speed the search for target images [26]. Although these methods achieve a great deal of success on many types of image the pre-defined feature measures and rules for applying them will preclude good search solutions in the general case.

The tracking of eye movements has been employed as a pointer and a replacement for a mouse [48], to vary the screen scrolling speed [47] and to assist disabled users [46]. However, this work has concentrated upon replacing and extending existing computer interface mechanisms rather than creating a new form of interaction. Indeed the imprecise nature of saccades and fixation points has prevented these approaches from yielding benefits over conventional human interfaces.

Notions of pre-attentive vision [25,32-34] and visual similarity are very closely related. Both aspects of human vision are relevant to content-based image retrieval; attention mechanisms tell us what is eye-catching and important within an image, and visual similarity tells us what parts of an image match a different image.

A more recent development has yielded a powerful similarity measure [35]. In this case the structure of a region in one image is being compared with random parts in a second image while seeking a match. This time if a *match* is found the score is increased, and a series of randomly generated features are applied to the *same* location in the second image that obtained the first match. A high scoring region in the second image is only reused while it continues to yield matches from randomly generated features and increases the similarity score. The conjecture that a region in the second image that shares a large number of different features with a region in the first image is perceptually similar is reasonable and appears to be the case in practice [35]. The measure has been tested on trademark images and fingerprints and within certain limits shown to be tolerant of translation, rotation, scale change, blur, additive noise and distortion. This approach does not make use of a pre-defined distance metric plus feature space in which feature values are extracted from a query image and used to match those from database

images, but instead generates features on a trial and error basis during the calculation of the similarity measure. This has the significant advantage that features that determine similarity can match whatever image property is important in a particular region whether it be a shape, a texture, a colour or a combination of all three. It means that effort is expended searching for the best feature for the region rather than expecting that a fixed feature set will perform optimally over the whole area of an image and over every image in the database. There are no necessary constraints on the pixel configurations used as features apart from the colour space and the size of the regions which is dependent in turn upon the definition of the original images.

More formally, in this method (full details of which are given in our European patent application 02252097.7), a first image (or other pattern) is represented by a first ordered set of elements A each having a value and a second pattern is represented by a second such set. A comparison of the two involves performing, for each of a plurality of elements $\underline{x}$ of the first ordered set the steps of selecting from the first ordered set a plurality of elements $\underline{x}$' in the vicinity of the element $\underline{x}$ under consideration, selecting an element $\underline{y}$ of the second ordered set and comparing the elements $\underline{x}$' of the first ordered set with elements $\underline{y}$' of the second ordered set (each of which has the same position relative to the selected element $\underline{y}$' of the second ordered set as a respective one $\underline{x}$' of the selected plurality of elements of the first ordered set has relative to the element $\underline{x}$ under consideration). The comparison itself comprises comparing the value of each of the selected plurality of elements $\underline{x}$' of the first set with the value of the correspondingly positioned element $\underline{y}$' of the like plurality of elements of the second set in accordance with a predetermined match criterion to produce a decision that the plurality of elements of the first ordered set matches the plurality of elements of the second ordered set. The comparison is them repeated with a fresh selection of the plurality of elements $\underline{x}$' of the first set and/or a fresh selection of an element $\underline{y}$ of the second ordered set generating a similarity measure V as a function of the number of matches. Preferably, following a comparison resulting in a match decision, the next comparison is performed with a fresh selection of the plurality of elements $\underline{x}$' of the first set and the same selection of an element $\underline{y}$ of the second set.

**Invention**

According to the present invention there is provided a method of retrieval of stored images stored with metadata for at least some of the stored images, the metadata comprising at least one entry specifying

(a) a part of the respective image;

(b) another stored image; and

(c) a measure of the degree of similarity between the specified part and the specified other stored image; the method comprising

i. displaying one or more images;

ii. receiving input from a user indicative of part of the displayed images;

iii. determining measures of interest for each of a plurality of non-displayed stored images specified by the metadata for the displayed image(s), as a function of the similarity measure(s) and the relationship between the user input and the part specified;

iv. selecting from those non-displayed stored images, on the basis of the determined measures, further images for display..

Other aspect of the invention are set out in the other claims.

**Examples**

Some embodiments of the invention will now be described, by way of example, with reference to the accompanying drawings, in which:

Figure 1 is a block diagram of an apparatus according to one embodiment of the invention; and

Figure 2 is a flowchart showing how that apparatus functions.

The apparatus shown in Figure 1 comprises a processor 1, a memory 3, disc store 4, keyboard 5, display 6, mouse 7, and telecommunications interface 8 such as might be found in a conventional desktop computer. In addition, the apparatus includes a gaze tracker 10, which is a system that observes, by means of a camera, the eye of a user and generates data indicating which part of the display 6 the user is looking at. One gaze tracker that might be used is the Eyegaze system, available from LC Technologies Inc., Fairfax, Virginia, U.S.A.. As well as the usual operating system software, the disc store 4 contains a computer program which serves to implement the method now to be described whereby the user is enabled to search a database of images. The database could be stored in the disc store 4, or it could be stored on a remote server accessible via the telecommunications interface 8.

**Basic Method**

The first method to be described, suitable for a small database, assumes that for each image stored in the database, the database already also contains one or more items of metadata each of which identifies a point or region of the image in question, another image, and a score indicating a degree of similarity between that point or region and the other image, For example a metadata item for an image frog.bmp might read:

113,42; toad.bmp;61

meaning that the image frog.bmp has, at x, y coordinates 113, 42, a feature which shows a similarity score of 61 with the image toad.bmp. Further such items might indicate similarities of some other location within frog.bmp to toad.bmp, or similarities between frog.bmp and further images in the database.

The manner in which such metadata can be created will be described later; first, we will describe a retrieval process, with reference to the flowchart of Figure 2.

The retrieval process begins at Step 1 with the display of some initial images from the database. These could be chosen (1a) by come conventional method (such as keywords) or (1b) at random. At Step 2 a "held image" counter is set to zero and, immediately the images are displayed, a timer defining a duration T is started (Step 3). During this time the user looks at the image and the system notes which of the images, and more particularly which parts of the images, the user finds to be of interest. This is done using the gaze tracker 10 which tracks the user's eye movement and records the position and duration of fixations (i.e. when the eye is not moving significantly). Its output takes the form of a sequence of reports each consisting of screen coordinates $x_s$, $y_s$ and the duration t of fixation at this point.

The value of T may be quite small allowing only a few saccades to take place during each iteration. This will mean that the displayed image set A will be updated frequently, but the content may not change dramatically at each iteration. On the other hand a large value of T may lead to most of the displayed images being replaced.

In Step 4, these screen coordinates are translated into an identifier for the image looked at, and x, y coordinates within that image. Also (Step 5) if there are multiple reports with the same x, y the durations t for these are added so that a single total duration $t_g$ is available for each x, y reported. Some users may suffer from short eye movements that do not provide useful information and so a threshold F may be applied so that any report with $t \leq F$ is discarded.

The next stage is to use this information in combination with metadata for the displayed images in order to identity images in the database which have similarities with those parts of the displayed images that the user has shown interest in.

Thus at Step 6, for a displayed image a and an image b in the database, a level of interest $I_{ab}$ is calculated. For this purpose the user is considered to have been looking at a particular point if the reported position of his gaze is at, or within, some region centred on, the point on question. The size of this region will depend on the size of the user's fovea centralis and his viewing distance from the screen: this may if desired be calibrated, though satisfactory results can be obtained if a fixed size is assumed.

For a displayed image a and an image b in the database, a level of interest $I_{ab}$ is calculated as follows:

$$I_{ab} = \sum_{g=1}^{G} \sum_{i=1}^{I} t_g . S_{abi} . \delta\left(x_g, y_g, x_i, y_i\right)$$

where $t_g$ is the total fixation duration at position $x_g$, $y_g$ ($g = 1, \ldots, G$) and G is the number of total durations. $S_{abi}$ is the score contained in the metadata for image a indicating a similarity between point $x_i$, $y_i$ in image a and another image b, and there are I items of metadata in respect of image a and specifying the same image b. Naturally, if, for any pair a, b, there is no metadata entry for $S_{abi}$, $S_{abi}$ is deemed to be zero. And $\delta(x_g, y_g, x_i, y_i)$ is 1 if $x_g$, $y_g$ is within the permitted region centred on $x_i$, $y_i$ and zero otherwise. For a circular area, $\delta = 1$ if and only if

$(x_g - x_i)^2 + (y_g - y_i)^2 < r^2$ where r is assumed effective radius of the fixation area. Obviously $I_{ab}$ exists only for those images b for which values of $S_{abi}$ are present in the metadata for one or more of the displayed images a.

The next (Step 7) is to obtain a score $I_b$ for such images, namely

$I_b = \sum I_{ab}$

summed over all the displayed images a.

Also in Step 7, the images with the highest values of $I_b$ are retrieved from the database and displayed. The number of images that are displayed may be fixed, or, as shown may depend on the number of images already held (see below).

Thus, if the number of images held is M and the number of images that are allowed to be displayed is N (assumed fixed) then the N-M highest scoring images will be chosen. The display is then updated by removing all the existing displayed images (other than held ones) and displaying the chosen images B instead. The images now displayed then become the new images A for a further iteration.

At Step 8 the user is given the option to hold any or all (thereby stopping the search) of the images currently displayed and prevent them from being overwritten in subsequent displays. The user is also free to release images previously held. The hold and release operations may be performed by a mouse click, for example. The value of M is correspondingly updated.

In Step 9 the user is able to bar displayed images from being subsequently included in set B and not being considered in the search from that point. It is common for image databases to contain many very similar images, some even being cropped versions of each other, and although these clusters may be near to a user's requirements, they should not be allowed to block a search from seeking better material. This operation may be carried out by means of a mouse click, for example.

The user is able to halt the search in Step 10 simply by holding all the images on the screen however, other mechanisms for stopping the search may be employed.

It should be noted that the user is able to invoke Steps 8 or 9 at any time in the process after Step 2. This could be a mouse click or a screen touch and may be carried out at the same time as continuing to gaze at the displayed images.

**Setting up the Database**

The invention does not presuppose the use of any particular method for generating the metadata for the images in the database. Indeed, it could in principle be generated manually. In general this will be practicable only for very small databases, though in some circumstances it may be desirable to generate manual entries in addition to automatically generated metadata.

We prefer to use the method described in our earlier patent application referred to above.

For a small database, it is possible to perform comparisons for every possible pair of images in the database, but for larger databases this is not practicable. For example if a database has 10,000 images this would require $10^8$ comparisons.

Thus, in an enhanced version, the images in the database are clustered; that is, certain images are designated as vantage images, and each cluster consists of a vantage image and a number of other images. It is assumed that this clustering is performed manually by the person loading images into the database. For example if he is to load a number of images of horses, he might choose one representative image as the vantage image and mark others as belonging to the cluster. Note that an image may if desired belong to more than one cluster.

The process of generating metadata is then facilitated:

(a) Each image in a cluster is scored against every other image in its own cluster (or clusters).

(b) Each vantage image is scored against every other vantage image.

The possibility however of other links being also generated is not excluded. In particular, once a database has been initially set up in this way one could if desired make further comparisons between images, possibly at random, to generate more metadata, so that as time goes on more and more links between images are established.

**External Images**

In the above-described retrieval method it was assumed that the initial images were retrieved at random or by some conventional retrieval method. A better option is to allow the user to input his own images to start the search (Step 1c). In this case, before retrieval can commence it is necessary to set up metadata for these external starting images. This is done by running the set-up method to compare (Step 1d) each of these starting images with all the images in the database (or, in a large database all the vantage images). In this way the starting

images (temporarily at least) effectively become part of the database and the method then proceeds in the manner previously described.

**Variations**

The "level of interest" is defined above as being formed from the products of the durations $t_g$ and the scores S; however other monotonic functions may be used. The set-up method (and hence also the retrieval method) described earlier assumes that a metadata entry refers to a particular point within the image. Alternatively, the scoring method might be modified to perform some clustering of points so that an item of metadata, instead of stating that a point (x, y) in A has a certain similarity to B, states that a region of specified size and shape, at (x, y) in A, has a certain similarity to B. One method of doing this, which assumes a square area of fixed size $2\Delta+1$ x $2\Delta+1$, is as follows. Starting with the point scores S(x, y):

- for each point, add the scores for all pixels with such an area centred on x, y to produce an area score $S^1(x,y)=\sum_{u=x-\Delta}^{x+\Delta}\sum_{v=y-\Delta}^{y+\Delta}S(u,v)$
- select one or more areas with the largest $S^1$.

Then $S^1$ are stored in the metadata instead of S. The retrieval method proceeds as before except that (apart from the use of $S^1$ rather than S) the function $\delta$ is redefined as being 1 whenever the gaze point $x_g$ $y_g$ falls within the square area or within a distance r of its boundary.

If areas of variable size and/or shape are to be permitted then naturally the metadata would include a definition of the size and shape and the function $\delta$ modified accordingly.

In the interests of avoiding delays, during Steps 2 to 6, all 'other' images referenced by the metadata of the displayed image could be retrieved from the database and cached locally.

Note that the use of a gaze tracker is not essential; user input by means of a pointing device such as a mouse could be used instead, though the gaze tracker option is considered to be much easier to use.

During the process of image retrieval users can traverse a sequence of images that are selected by the user from those presented by the computer. The machine endeavours to predict the most relevant groups of images and the user selects on the basis of recognised associations with a real or imagined target image. The retrieval will be successful if the images presented to the user are on the basis of the same associations that the user also recognises. Such associations might depend upon semantic or visual factors which can take virtually unlimited forms often dependent upon the individual user's previous experience and interests. This system makes provision for the incorporation of semantic links between images derived from existing or manually captured textual metadata.

The process of determining the similarity score between two images necessarily identifies a correspondence between regions that give rise to large contributions towards the overall image similarity. A set of links between image locations together with values of their strengths is then available to a subsequent search through images that are linked in this way. There may be several such links between regions in pairs of images, and further multiple links to regions in other images in the database. This network of associations is more general than those used in other content-based image retrieval systems which commonly impose a tree structure on the data, and cluster images on the basis of symmetrical distance measures between images [27,37]. Such restrictions prevent associations between images being offered to users that are not already present in the fixed hierarchy of clusters. It should be noted that the links in this system are not symmetric as there is no necessary reason for a region that is linked to a second to be linked in the reverse direction. The region in the second image may be more similar to a different region in the first image. The triangle inequality is not valid as it is quite possible for image A to be very similar to B, and B to C, but A can be very different from C. Other approaches preclude solutions by imposing metrics that are symmetric and/or satisfy the triangle inequality [28].

This new approach to content-based image retrieval will allow a large number of pre-computed similarity associations between regions within different images to be incorporated into a novel image retrieval system. In large databases it will not be possible to compare all images with each other so clusters and vantage images [37,38,39] will be employed to minimise computational demands. However, as users traverse the database fresh links will be continually generated and stored that may be used for subsequent searches and reduce the reliance upon vantage images. The architecture will be capable of incorporating extra links derived from semantic information [12] that already exists or which can be captured manually.

It is not natural to use a keyboard or a mouse when carrying out purely visual tasks and presents a barrier to many users. Eyetracking technology has now reached a level of performance that can be considered as an interface for image retrieval that is intuitive and rapid. If it is assumed that users fixate on image regions that attract their interest, this information may be used to provide a series of similar images that will converge upon the target or an image that meets the users' demands. Of course a mouse could be used for the same task, but has less potential for extremely rapid and intuitive access. Users would be free to browse in an open-ended manner or to seek a target image by just gazing at images and gaining impressions, but in so doing driving the search by means of saccades and fixation points. Similarity links between image regions together with corresponding strength values would provide the necessary framework for such a system which would be the first of its kind in the world.

**References**

[1] A.W.M. Smeulders, M. Worring, S. Santini, A. Gupta and R. Jain, "Content-Based Image Retrieval at the End of the Early Years," IEEE Trans PAMI, Vol 22, No 12, pp 1349-1379, December 2000.

[2] Y. Rui, T.S. Huang and S-F Chang, "Image Retrieval: Current Techniques, Promising Directions and Open Issues,"

[3] S-K. Chang and A. Hsu, "Image Information Systems: Where Do We Go from Here?" IEEE Trans on Data and Knowledge Eng., Vol 4, No 5, pp 431-442, October 1992.

[4] M. Flickner, H. Sawhney, W. Niblack, J. Ashley, Q. Huang, B. Dom, M. Gorkani, J. Hafner, D. Lee, D. Petkovic, D. Steele, and P. Yanker, "Query by Image and Video Content: The QBIC System," IEEE Computer, 1995.

[5] A. Pentland, R. W. Pickard, and S. Sclaroff, "Photobook: Content-Based Manipulation of Image Databases," SPIE Storage and Retrieval of Images and Video Databases II, No 2185, Feb 6-10, San Jose, 1994.

[6] C. Carson, S. Belongie, H. Greenspan, and J. Malik, "Blobworld: Image Segmentation using Expectation-Maximisation and its Application to Image Querying," IEEE Trans. Pattern Analysis and Machine Intelligence, Vol 24, No 8, pp 1026-1038, August 2002.

[7] J. R. Smith and S-F. Chang, "VisualSEEk: a fully automated Content-Based Image Query System," Proc. ACM Int. Conf. Multimedia, pp 87-98, Boston MA, November 1996.

[8] W-Y. Ma and B. S. Manjunath, "NeTra: a Toolbox for Navigating Large Image Databases," Multimedia Systems, Vol 7, pp 184-198, 1999.

[9] A. Gupta and R. Jain, "Visual Information Retrieval," Communications of the ACM, Vol 40, No 5, pp 70-79, May 1997.

[10] J. Dowe, "Content Based Retrieval in Multimedia Imaging," Proc SPIE Conf. Storage and Retrieval for Image and Video Databases, 1993.

[11] J. R. Smith and S-F. Chang, "Integrated Spatial and Feature Image Query," Multimedia Systems, Vol 7, No 2, pp 129-140, 1999.

[12] I. J. Cox, M. L. Miller, T. P. Minka, and T. V. Papathomas, "The Baysian Image Retrieval System, PicHunter: Theory, Implementation, and Psychophysical Experiments," IEEE Trans. Image Processing, Vol 9, No 1, pp20-37, 2000.

[13] J. Matas, D. Koubaroulis, and J. Kittler, "Colour Image Retrieval and Object Recognition using the Multimodal Neighbourhood Signature," Proc. ECCV, LNCS, Vol 1842, pp 48-64, Berlin, June 2000.

[14] D. Koubaroulis, J. Matas, and J. Kittler, "Evaluating Colour-Based Object Recognition Algorithms Using the SOIL-47 Database," 5th Asian Conf. on Computer Vision, Melbourne, Australia, 23-25 January 2002.

[15] M Westmacott, P. Lewis, and Kirk Martinez, "Using Colour Pair Patches for Image Retrieval," Proc. 1st European Conf. on Colour in Graphics, Image and Vision, pp 245-247, April 2002.

[16] G. Pass, R. Zabih, and J. Miller, "Comparing Images using Color Coherence Vectors," 4th ACM Conf. on Multimedia, Boston MA, November 1996.

[17] http://www.artisteweb.org/

[18] S. Chan, K. Martinez, P. Lewis, C. Lahanier, and J. Stevenson, "Handling Sub-Image Queries in Content-Based Retrieval of High Resolution Art Images," International Cultural Heritage Informatics Meeting 2, pp 157-163, September 2001.

[19] D. Dupplaw, P. Lewis, and M. Dobie, "Spatial Colour Matching for Content Based Retrieval and Navigation," Challenge of Image Retrieval ,99, Newcastle, February 1999.

[20] P. Allen, M. Boniface, P. Lewis, and K. Martinez, "Interoperability between Multimedia Collections for Content and Metadata-Based Searching," Proc. WWW Conf., May 2002.

[21] T. Louchnikova and S. Marchand-Maillet, "Flexible Image Decomposition for Multimedia Indexing and Retrieval," Proc SPIE Internet Imaging III, Vol 4673, 2002.

[22] J. Fan, M. Body, X. Zhu, M-S. Hacid, and E. El-Kwae, "Seeded Image Segmentation for Content-Based Image Retrieval Application," Proc SPIE, Storage and Retrieval for Media Databases, 2002.

[23] J. Puzicha, T. Hofmann, and J. M. Buhmann, "Histogram Clustering for Unsupervised Segmentation and Image Retrieval," Pattern Recognition Letters, 20, pp 899-909, 1999.

[24] A. Mojsilovic, J. Kovacevic, J. Hu, R. J. Safranek, and S. K. Ganapathy, "Matching and Retrieval Based on the Vocabulary and Grammar of Color Patterns," IEEE Trans on Image Processing, Vol 9, No 1, January 2000.

[25] E. Niebur, L. Itti, and Christof Koch, "Controlling the Focus of Visual Selective Attention," in Models of Neural Networks IV, Van Hemmen, Cowan & Domany eds., Springer Verlag, NY, pp 247-276, 2002.

[26] H. D. Tagare, K Toyama, and J. G. Wang, "A Maximum-Likelihood Strategy for Directing Attention during Visual Search," IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol 23, No 5, May 2001.

[27] S. Berretti, A. D. Bimbo, and E. Vicario, "Efficient Matching and Indexing of Graph Models in Content-Based Retrieval," IEEE Trans on Pattern Analysis and Machine Intelligence, Vol 23, No 10, October 2001.

[28] S. Santini and R. Jain, "Similarity Measures,", IEEE Trans on Pattern Analysis and Machine Intelligence, Vol 21, No 9, September 1999.

[29] A. Vailaya, M. A. T. Figueiredo, A. K. Jain, and H-J. Zhang, "Image Classification for Content-Based Indexing," IEEE Trans on Image Processing, Vol 10, No 1, pp 117-130, January 2001.

[30] J. Z. Wang, J. Li, and G. Wiederhold, "SIMPLIcity: Semantics-Sensitive Integrated Matching for Picture LIbraries," IEEE Trans on Pattern Analysis and Machine Intelligence, Vol 23, No 9, pp 947-963, September 2001.

[31] Y. Rui, T. S. Huang, M. Ortega, and S. Mehrotra, "Relevance Feedback: A Power Tool for Interactive Content-Based Image Retrieval," IEEE Trans on Circuits and Video Technology, pp 1-13, 1998.

[32] F. W. M. Stentiford, "An evolutionary programming approach to the simulation of visual attention," Congress on Evolutionary Computation, Seoul, May 27-30, 2001.

[33] F. W. M. Stentiford, "An estimator for visual attention through competitive novelty with application to image compression," Picture Coding Symposium, Seoul, 24-27 April, 2001.

[34] F. W. M. Stentiford, N. Morley, and A. Curnow, "Automatic identification of regions of interest with application to the quantification of DNA damage in cells," *Proc SPIE,* Vol 4662, San Jose, 20-26 Jan, 2002.

[35] F. W. M. Stentiford, "An attention based similarity measure with application to content based information retrieval," accepted to the Storage and Retrieval for Media Databases conference at SPIE Electronic Imaging 2003, 20-24 Jan, Santa Clara.

[36] TriTex Project IST-1999-20500, "Automated 3D Texture Content Management in Large Scale Data Sets," http://www.connect.slb.com/Docs/ofs/ofs_research_public/ofsr_public_images/locations/ssr/stavanger/TriTex/

[37] J. Landre and F. Truchetet, "A hierarchical architecture for content-based image retrieval of paleontology images," *Proc SPIE,* Vol 4676, San Jose, 20-26 Jan, 2002.

[38] P. N. Yianilos, "Data structures and algorithms for nearest neighbor search in general metric spaces," Proc. 4th ACM-SIAM Symposium on Discrete Algorithms, pp 311-321, 1993.

[39] J.Vleugels and R. Veltkamp, "Efficient image retrieval through vantage objects", Pattern Recognition, Vol 35, pp 69-80, 2002.

[40] A. B. Benitez, M. Beigi, and S-F. Chang, "Using relevance feedback in content-based image metasearch," IEEE Internet Computing, pp 59-69, July/August 1998.

[41] G. Ciocca and R. Schettini, "A multimedia search engine with relevance feedback," *Proc SPIE,* Vol 4672, San Jose, 20-26 Jan, 2002.

[42] L. Taycher, M. La Cascia, and S. Sclaroff, "Image digestion and relevance feedback in the ImageRover WWW search engine," Proc. 2nd Int. Conf. on Visual Information Systems, San Diego, pp 85-94, December 1997.

[43] J. Vendrig, M. Worring, and A. W. M. Smeulders, "Filter image browsing: exploiting interaction in image retrieval," Proc. 3rd Int. Conf. VISUAL'99, June 1999.

[44] R. B. Fisher and A. MacKirdy, "Integrated iconic and structured matching," Proc 5th European Conf. on Computer Vision, Vol II, pp 687-698, Frieburg, June 1998.

[45] P. R. Hill, C. N. Canagarajah, and D. R. Bull, "Texture gradient based watershed segmentation," ICASSP, Orlando, May 13-17, pp 3381-3384, 2002.

[46] F. Corno, L. Farinetti and I. Signorile, "A cost effective solution for eye-gaze assistive technology," 2002 IEEE Int Conf. on Multimedia and Expo, August 26-29, Lausanne, 2002.

[47] T. Numajiri, A. Nakamura, and Y. Kuno, "Speed browser controlled by eye movements," 2002 IEEE Int Conf. on Multimedia and Expo, August 26-29, Lausanne, 2002.

[48] J. P. Hansen, A. W. Anderson, and P. Roed, "Eye gaze control of multimedia systems," Symbiosis of Human and Artifact (Y. Anzai, K. Ogawa, and H. Mori (eds), Vol 20A, Elsevier Science, pp 37-42, 1995.

# CLAIMS

1. A method of retrieval of stored images stored with metadata for at least some of the stored images, the metadata comprising at least one entry specifying

(a) a part of the respective image;

(b) another stored image; and

(c) a measure of the degree of similarity between the specified part and the specified other stored image; the method comprising

i. displaying one or more images;

ii. receiving input from a user indicative of part of the displayed images;

iii. determining measures of interest for each of a plurality of non-displayed stored images specified by the metadata for the displayed image(s), as a function of the similarity measure(s) and the relationship between the user input and the part specified;

iv. selecting from those non-displayed stored images, on the basis of the determined measures, further images for display.

2. A method according to Claim 1 in which the receiving of input from a user is performed by means operable to observe movement of the user's eye.

3. A method according to claim 1 or 2 in which the user input identifies image locations and associated attention durations, and each measure of interest is the sum of individual measures for each identified location that is within a predetermined distance of a specified part, each said individual measure being a function of the attention duration that is associated with the identified location and the similarity measure that is associated with the specified part.

4. A method according to claim 3 in which each individual measure is the product of the duration and the similarity measure.

5. A method according to any one of the preceding claims in which the specified parts of the images are points within the images.

6. A method according to any one of claims 1 to 4 in which the specified parts of the images are regions thereof.

7. A method according to any one of the preceding claims in which steps (ii) to (iv) are repeated at least once.

8. A method according to any one of the preceding claims further including the initial steps of:

receiving one or more external images;

generating said metadata in respect of the external image(s); and

displaying the external image(s).

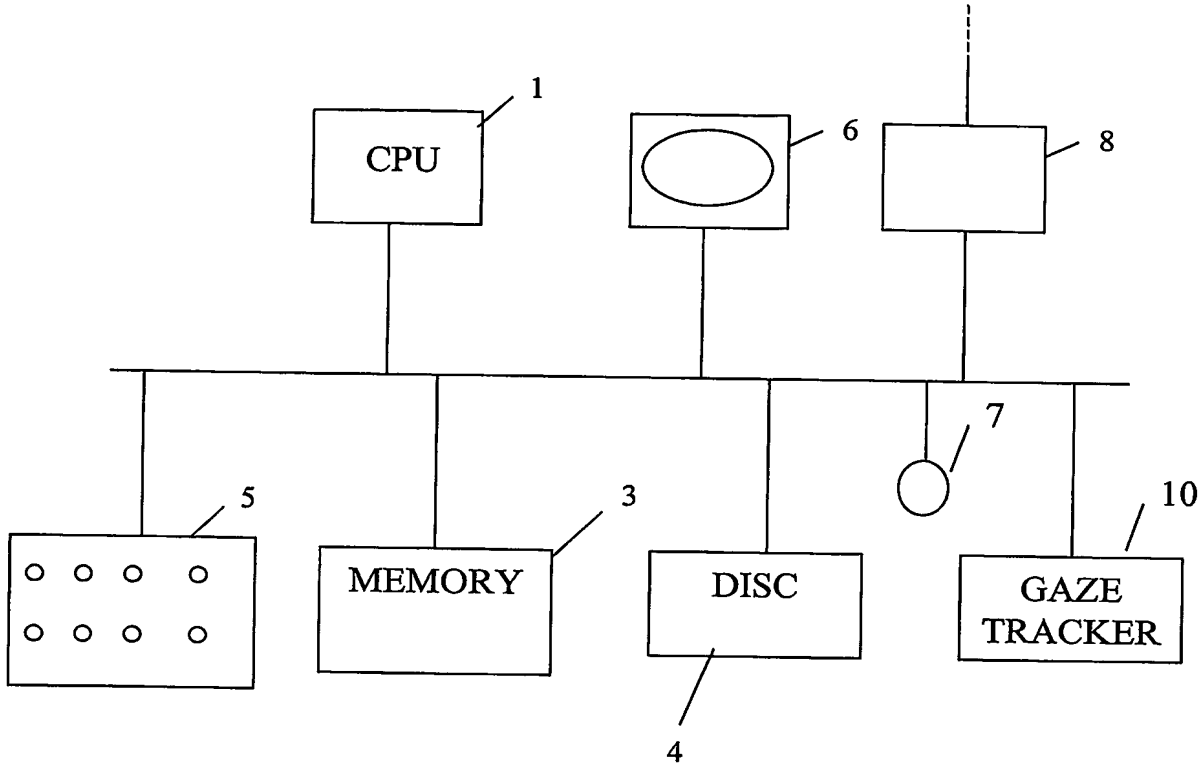



**Figure 1**

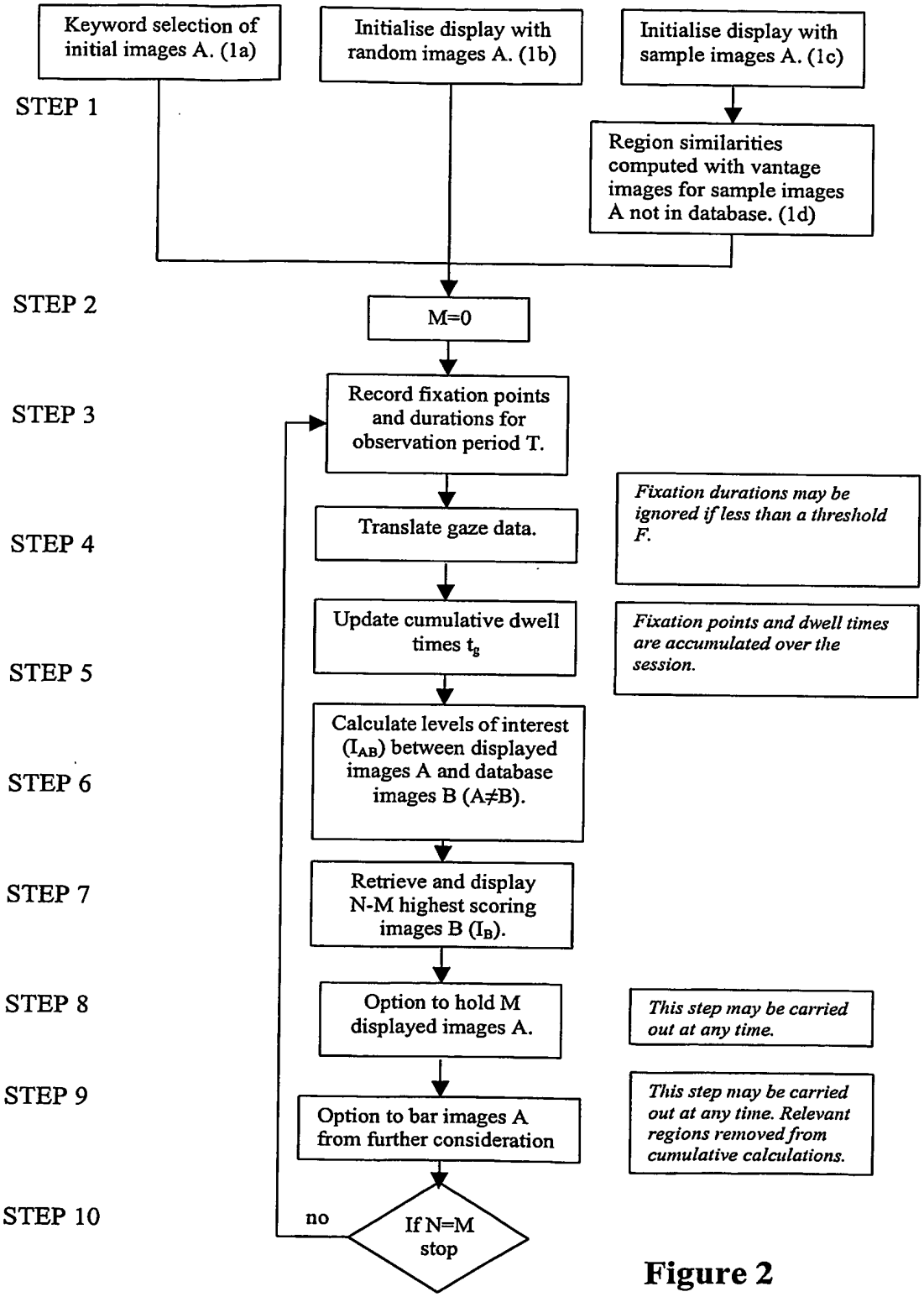

Keyword selection of initial images A. (1a)
Initialise display with random images A. (1b)
Initialise display with sample images A. (1c)
STEP 1
Region similarities computed with vantage images for sample images A not in database. (1d)
STEP 2
M=0
STEP 3
Record fixation points and durations for observation period T.
STEP 4
Translate gaze data.
Fixation durations may be ignored if less than a threshold F.
STEP 5
Update cumulative dwell times $t_g$
Fixation points and dwell times are accumulated over the session.
STEP 6
Calculate levels of interest ($I_{AB}$) between displayed images A and database images B ($A \neq B$).
STEP 7
Retrieve and display N-M highest scoring images B ($I_B$).
STEP 8
Option to hold M displayed images A.
This step may be carried out at any time.
STEP 9
Option to bar images A from further consideration
This step may be carried out at any time. Relevant regions removed from cumulative calculations.
STEP 10
no
If N=M stop

**Figure 2**